LENDING CLUB CASE STUDY

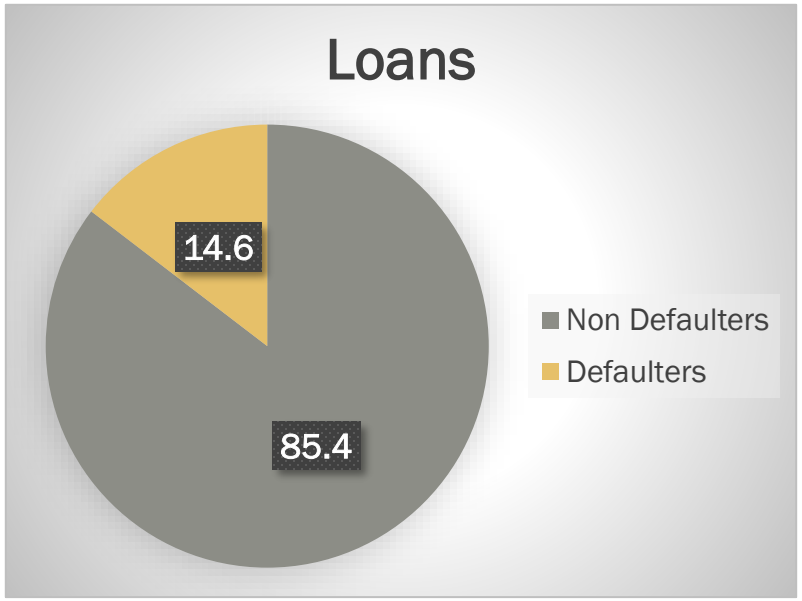
Data Understanding

Data Cleaning

Data Analysis and Recommendations

Data Understanding

Rows - 39717
Columns - 111



Data Cleaning

- **Removing the NULL values**
  - There are 57 columns with more than 50% NULL values.
- **Removing Columns with zero variance**
  - There are 6 columns with zero variance
- **Removing Behaviour variables which are of no use**
  - There are 21 such variables
- **Removing columns having only unique values**
  - There are two unique variables (id and member id)
- **Removing Rows having loan_status as “current”**

- **Cleaning interest rate column**
  - Removing ‘%’ sign from column values
- **Cleaning term column**
  - Removing the string ‘months’ from term column
- **Cleaning employment length column**
  - Removing ‘+’ sign from employment length column
- **Splitting the issue_d column into columns ‘month’ and ‘year’**

After removing all the irrelevant columns and rows from the dataset we have

- **38577 rows**
- **23 columns**

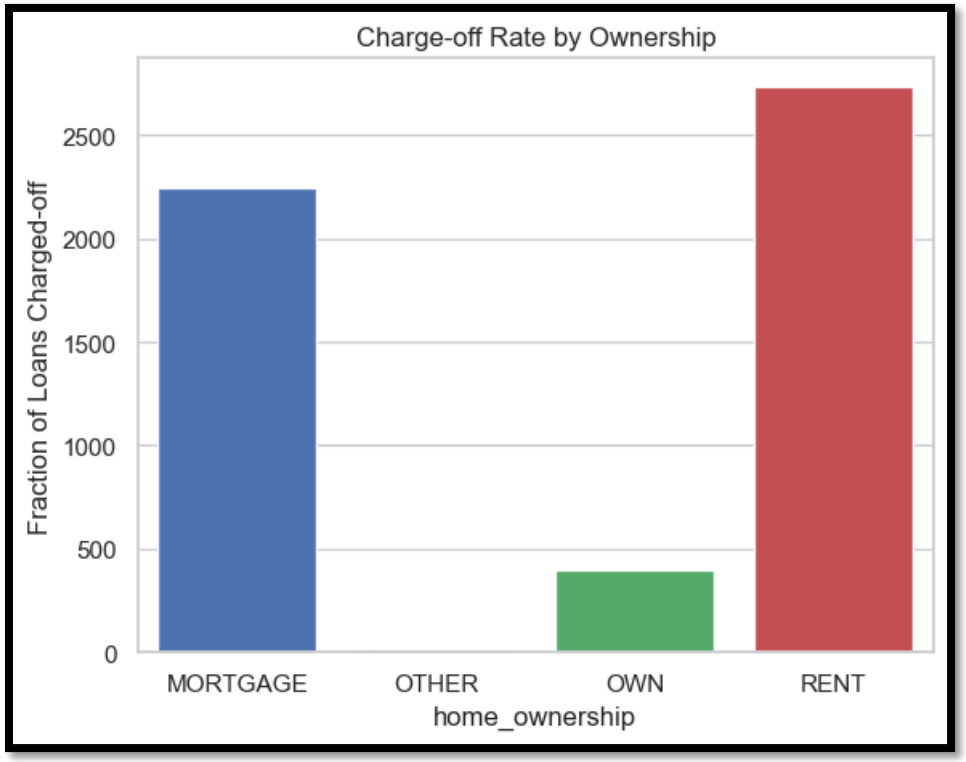
Variables of Interest

- 'loan_amnt',
- 'funded_amnt',
- 'funded_amnt_inv',
- 'term', 'int_rate',
- 'installment',
- 'grade',
- 'sub_grade',
- 'emp_title',
- 'emp_length',
- 'home_ownership',
- 'annual_inc',
- 'verification_status',
- 'issue_d'
- , 'loan_status',
- 'pymnt_plan',
- 'desc',
- 'purpose',
- 'dti',
- 'initial_list_status',
- 'pub_rec_bankruptcies',
- 'month',
- 'year'

Data Analysis
&
Recommendations

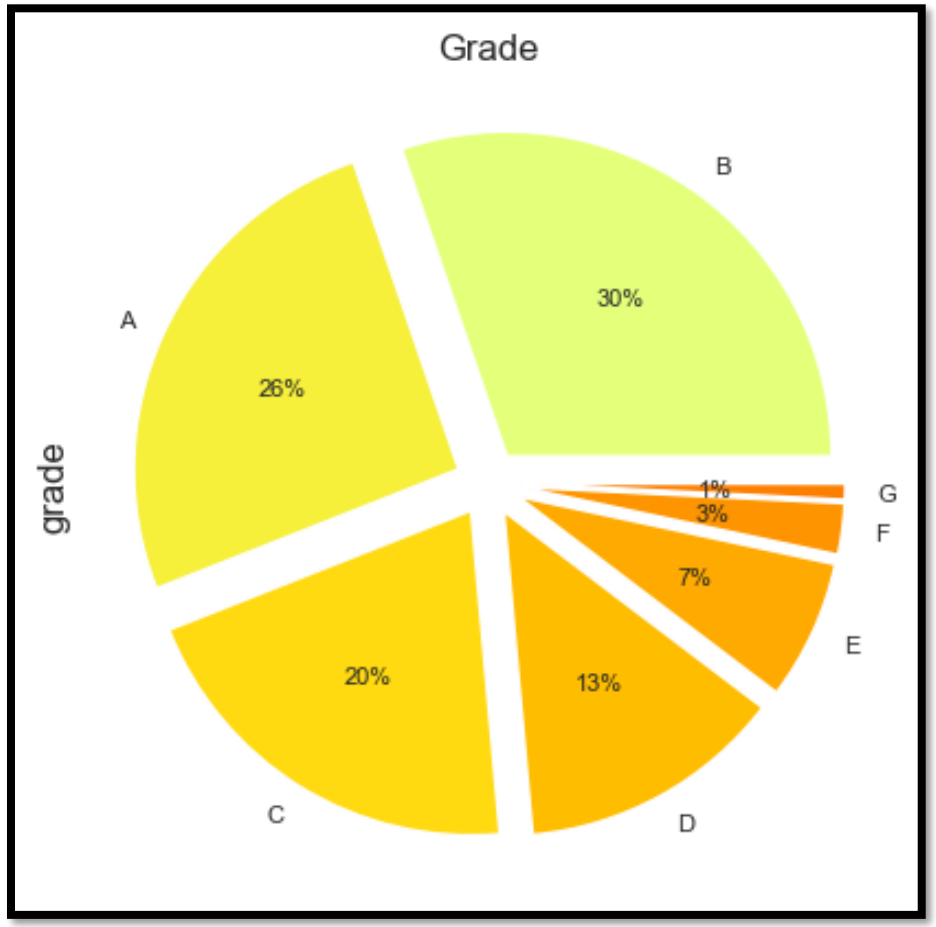
UNIVARIATE ANALYSIS

Home Ownership



Most of the applicants are either having rented houses or their home mortgaged.

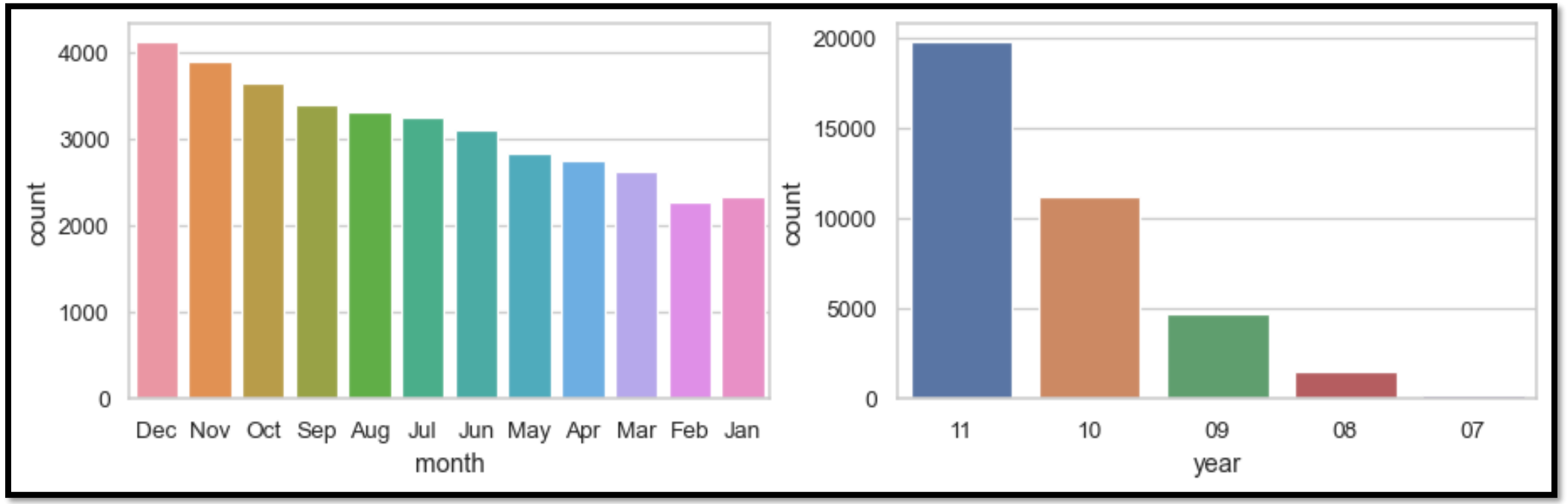
Grade Distribution



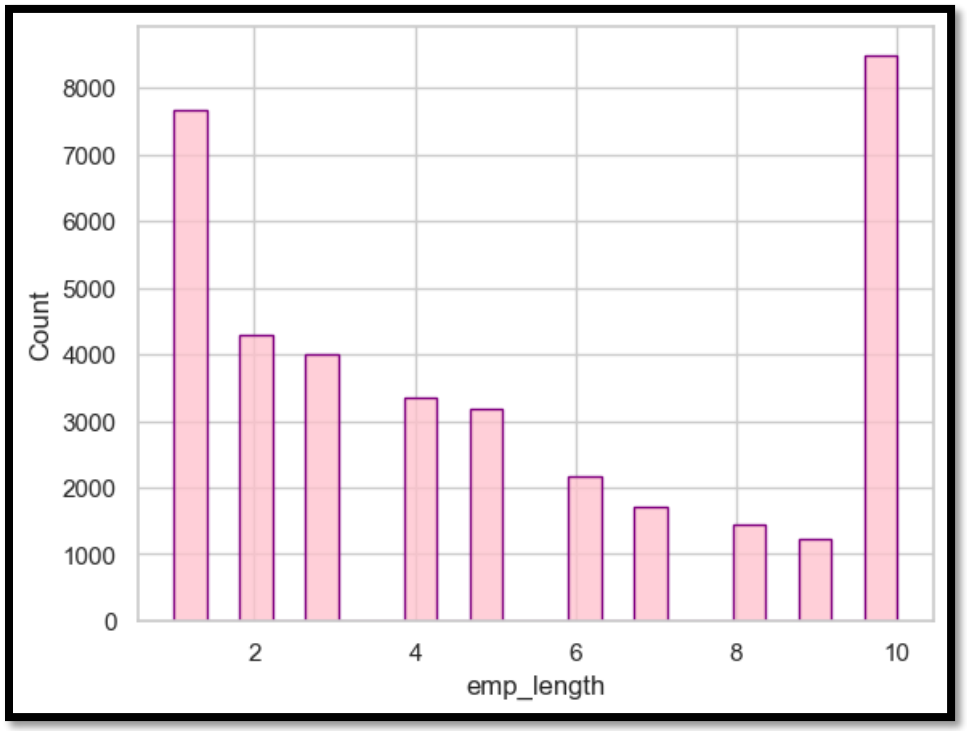
- 30% of the applicants fall in grade A and 26% falls in grade B.
- Only 1% of the applicants falls in grade G.

Issue Date Distribution

- More number of loans are issued in the month of December whereas least number of loans are issued in the month of February.
- More number of loans are issued in the year 2011 (53.2%) whereas least number of loans are issued in the year 2007 (0.7%).



Employment Length

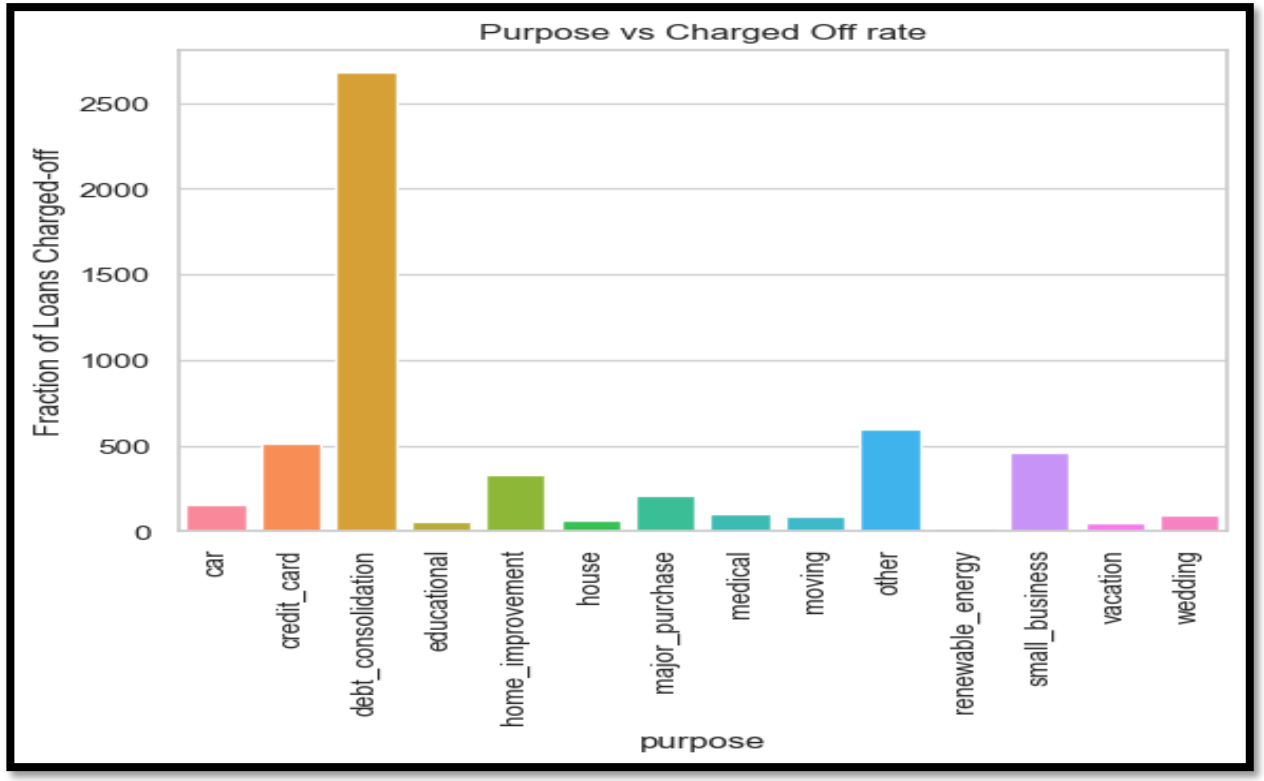


- Most of the employees are employed for 10 years or more.
- Comparatively a smaller number of employees with 9 years.

Purpose distribution

Loans issued for the purpose of debt consolidation is higher.

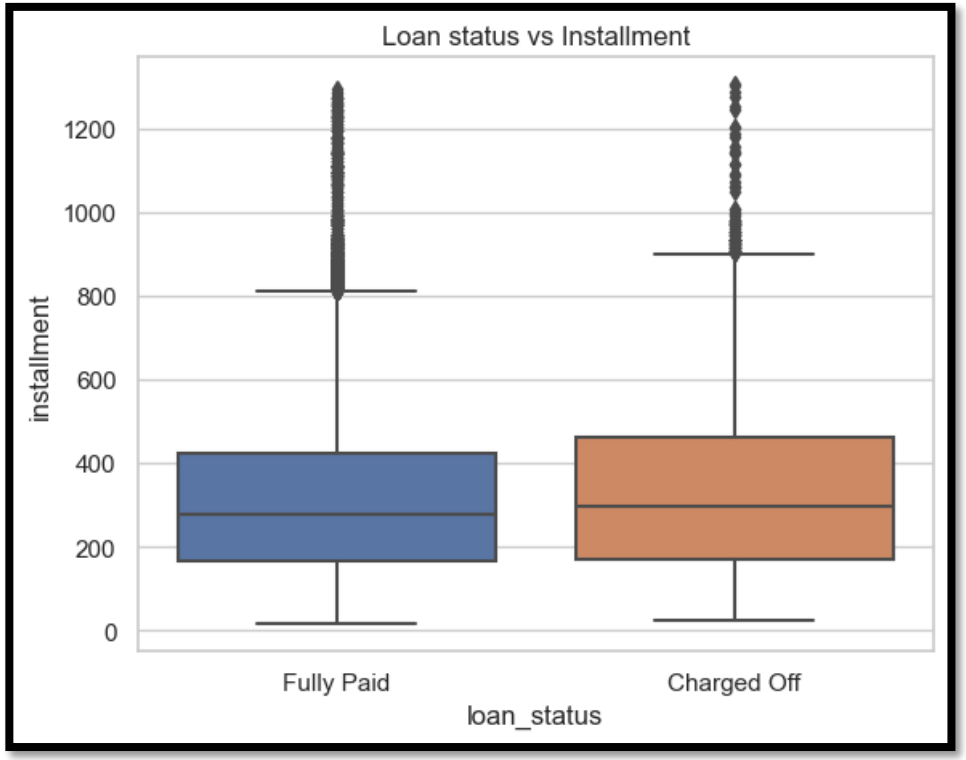
- ✓ 46.8% debt consolidation loans
- ✓ 13% credit card loans
- ✓ 7.5% home improvement loans
- ✓ 5.6% major purchase loans



Bivariate Analysis

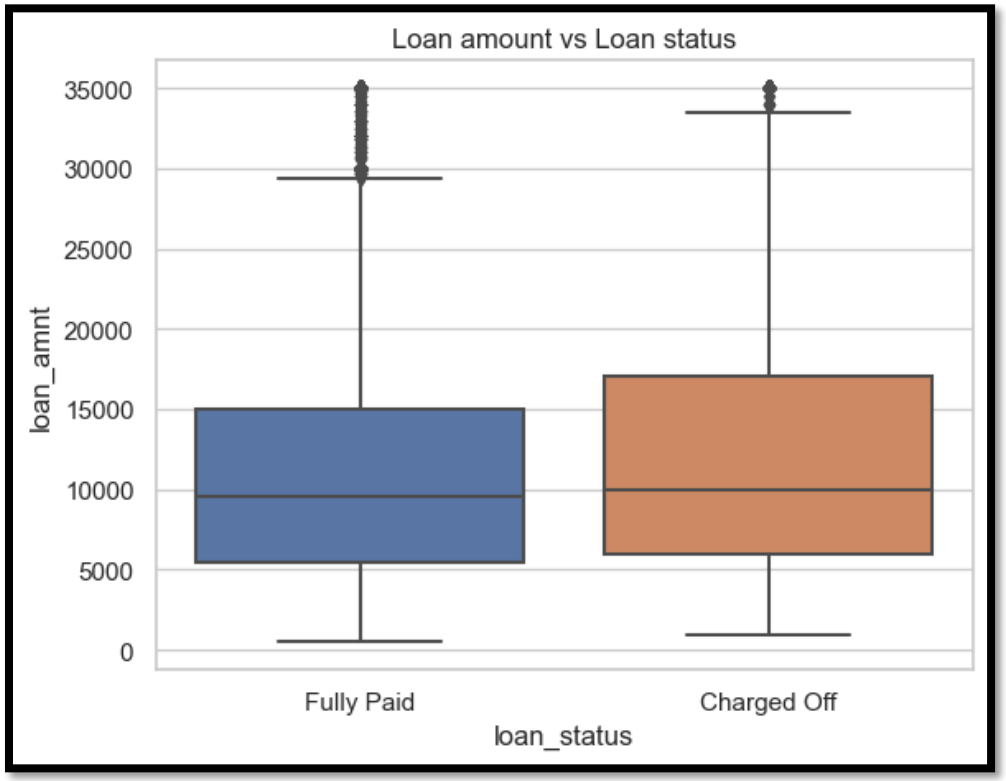
Installment

Charged-off loans tend to have higher installments.



Loan Status	mean	std	min	50%	max
0	322.09	207.38	15.69	277.70	1,295.21
1	339.87	217.11	22.79	297.53	1,305.19

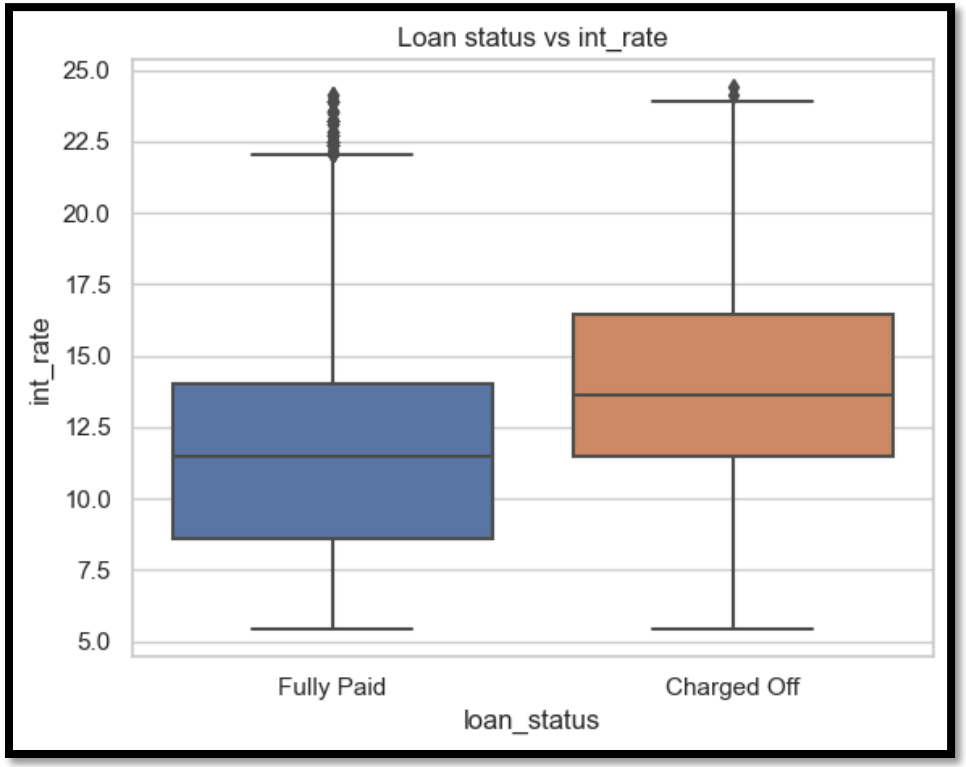
Loan Amount Distribution



- Charged-off loans tend to have higher loan amounts.
- Mean of Charged Off loan is higher than the fully paid loan.

Interest rate

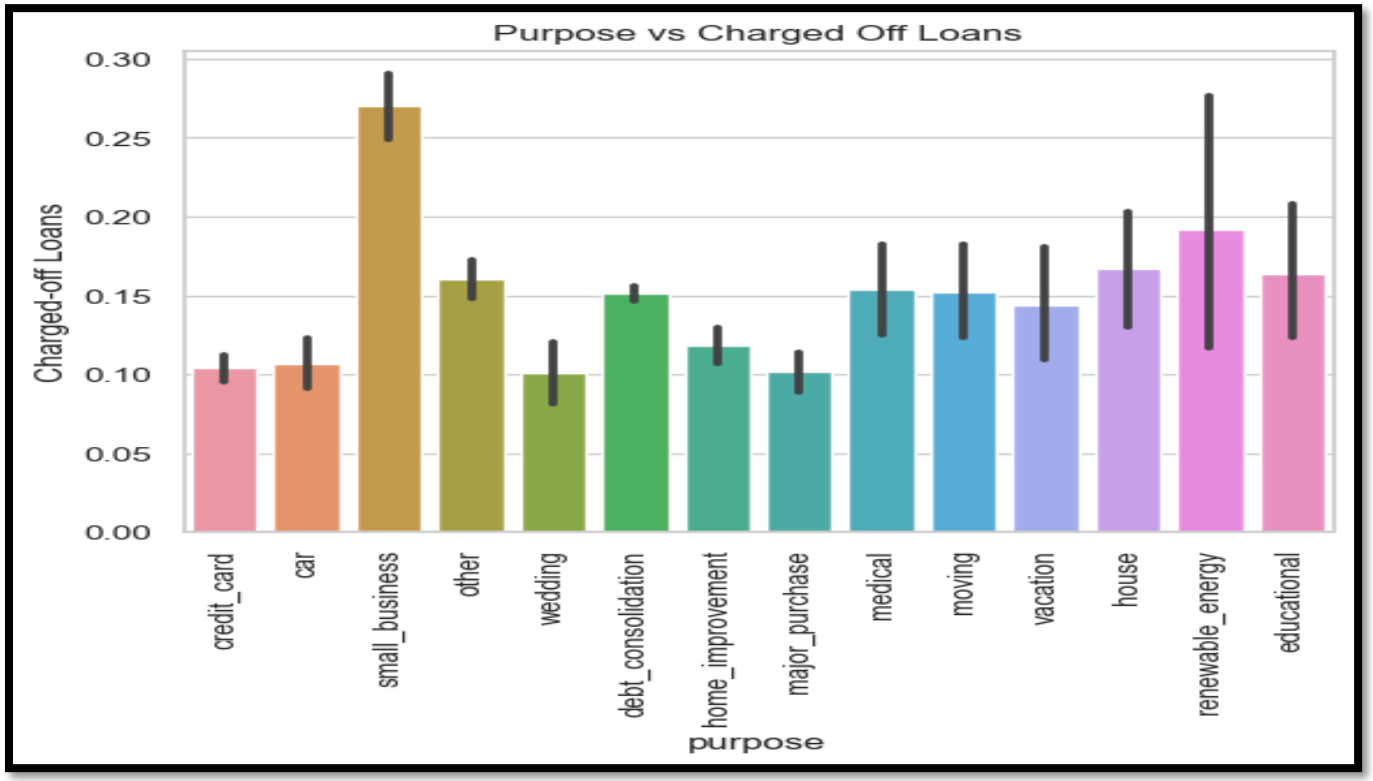
Charged-off loans tend to have much higher interest rates.



Purpose distribution- Defaulters

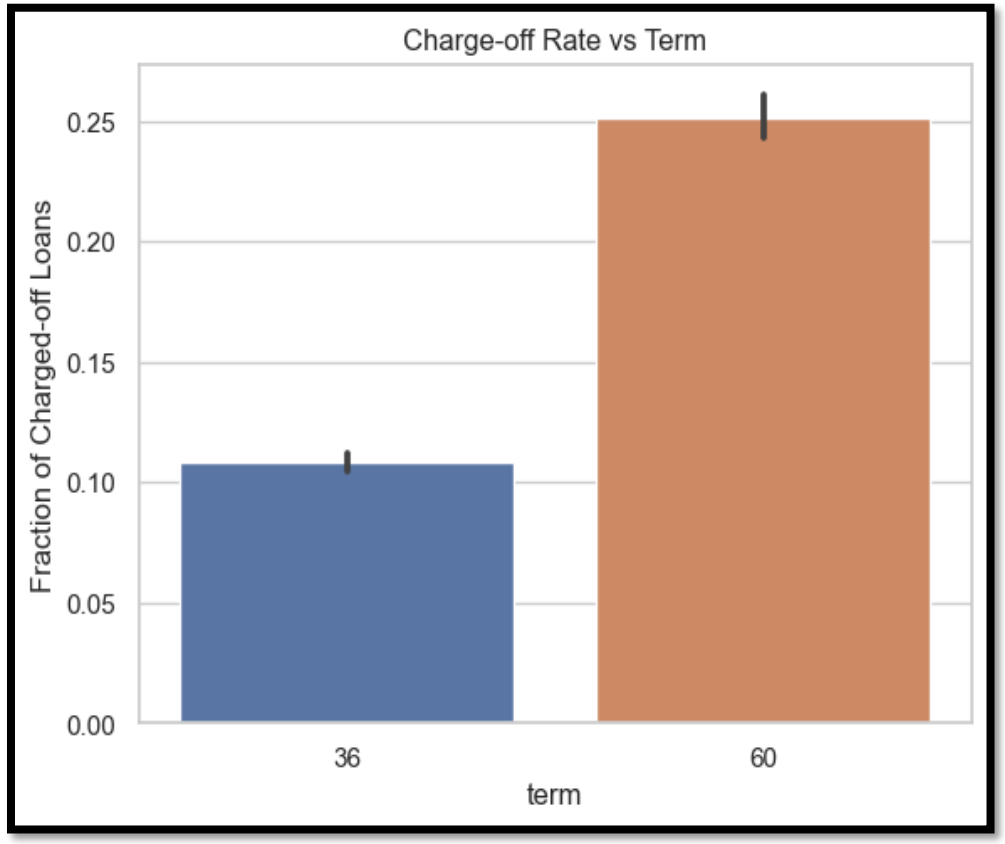
Defaulters are more for the purpose of small_businesses.

- Small business: 27%
- House: 17%
- Educational: 16%
- Debt consolidation: 15%
- Credit card: 10%
- Major purchase: 10%

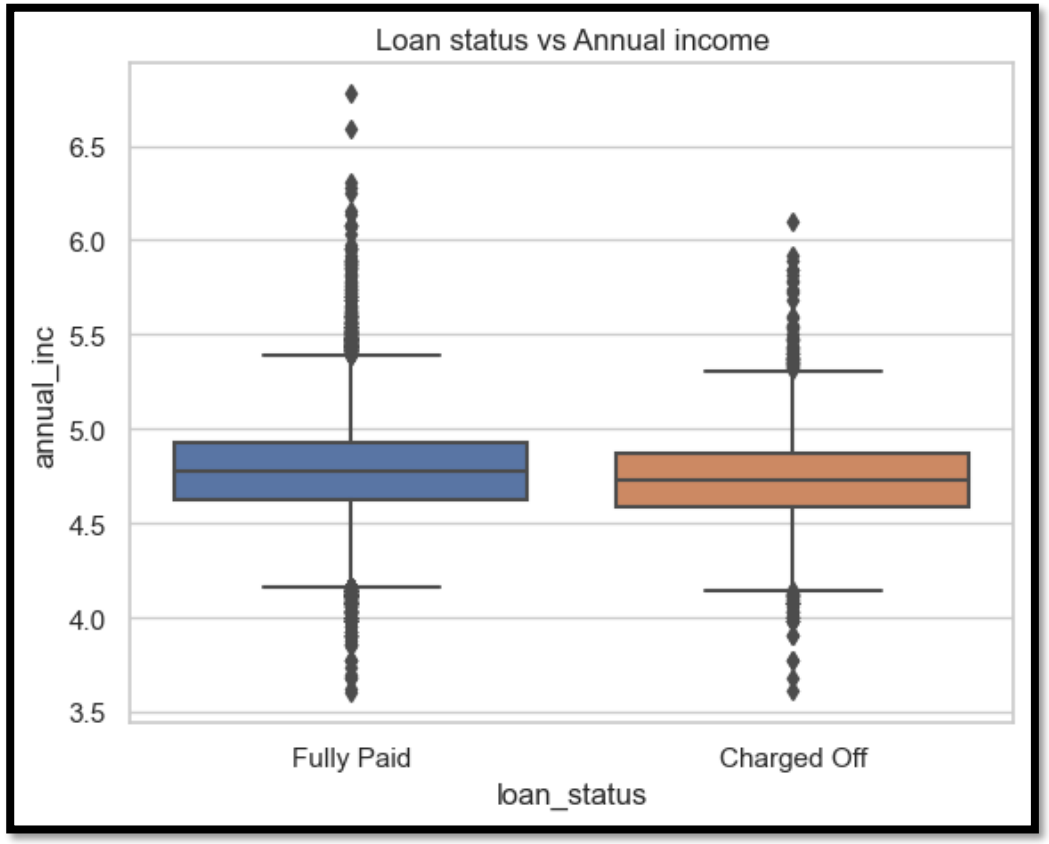


Term & Annual Income

Defaulters are more for the term 60 months

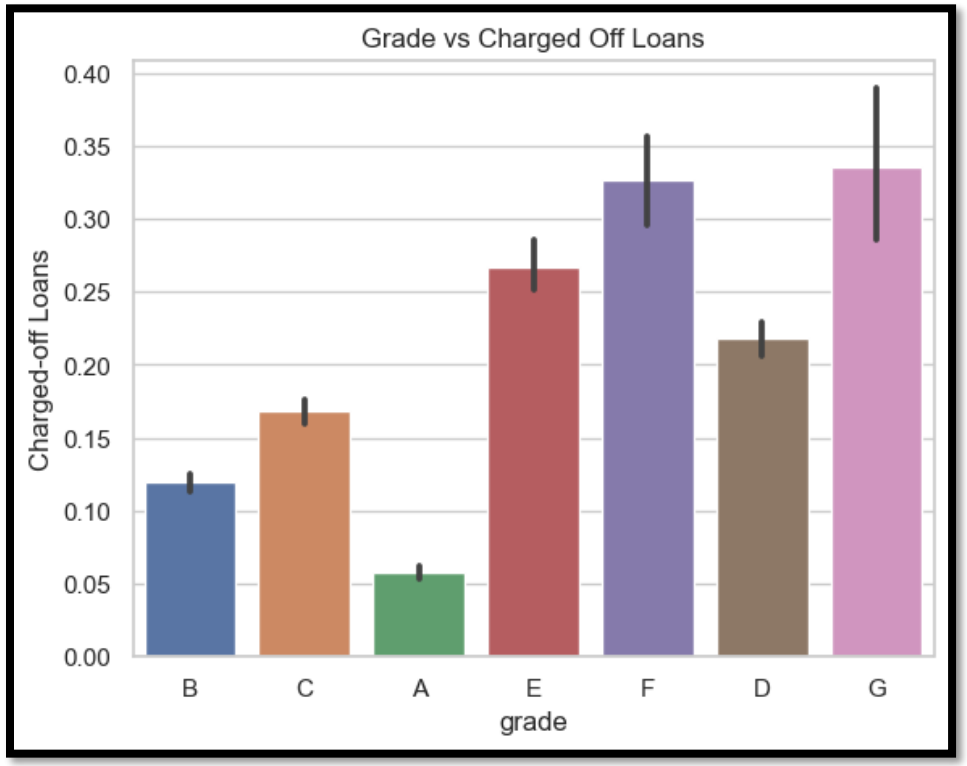


Those who earn more have fully paid their loans and defaulters have low income.

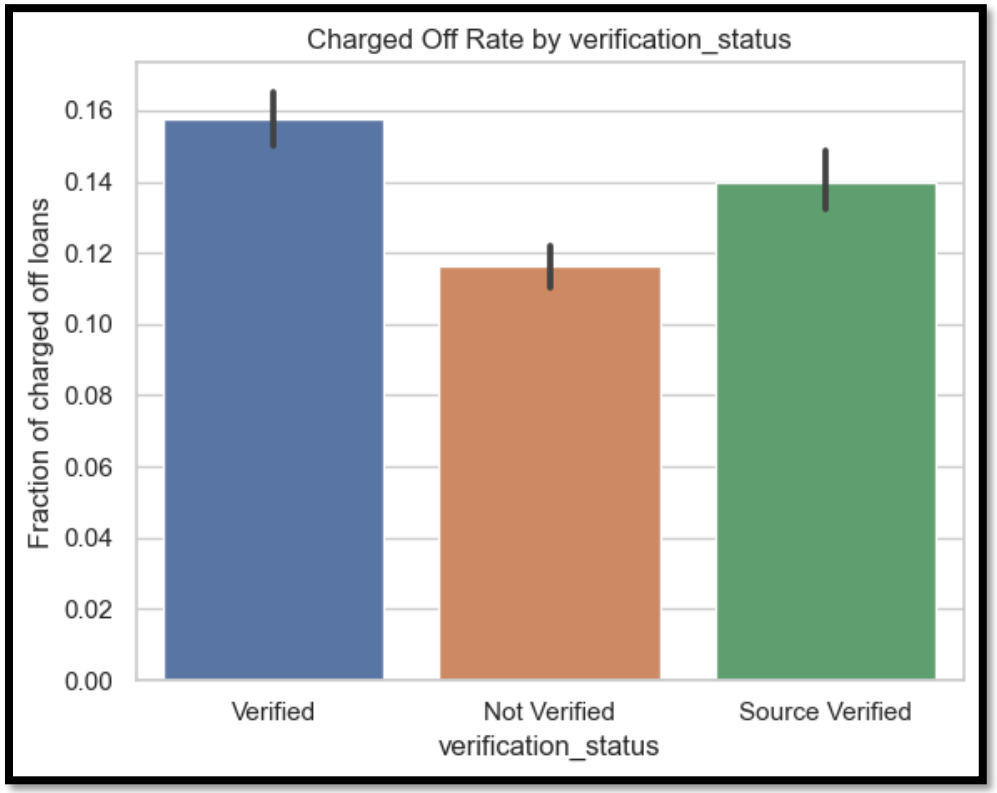


Grade & verification status

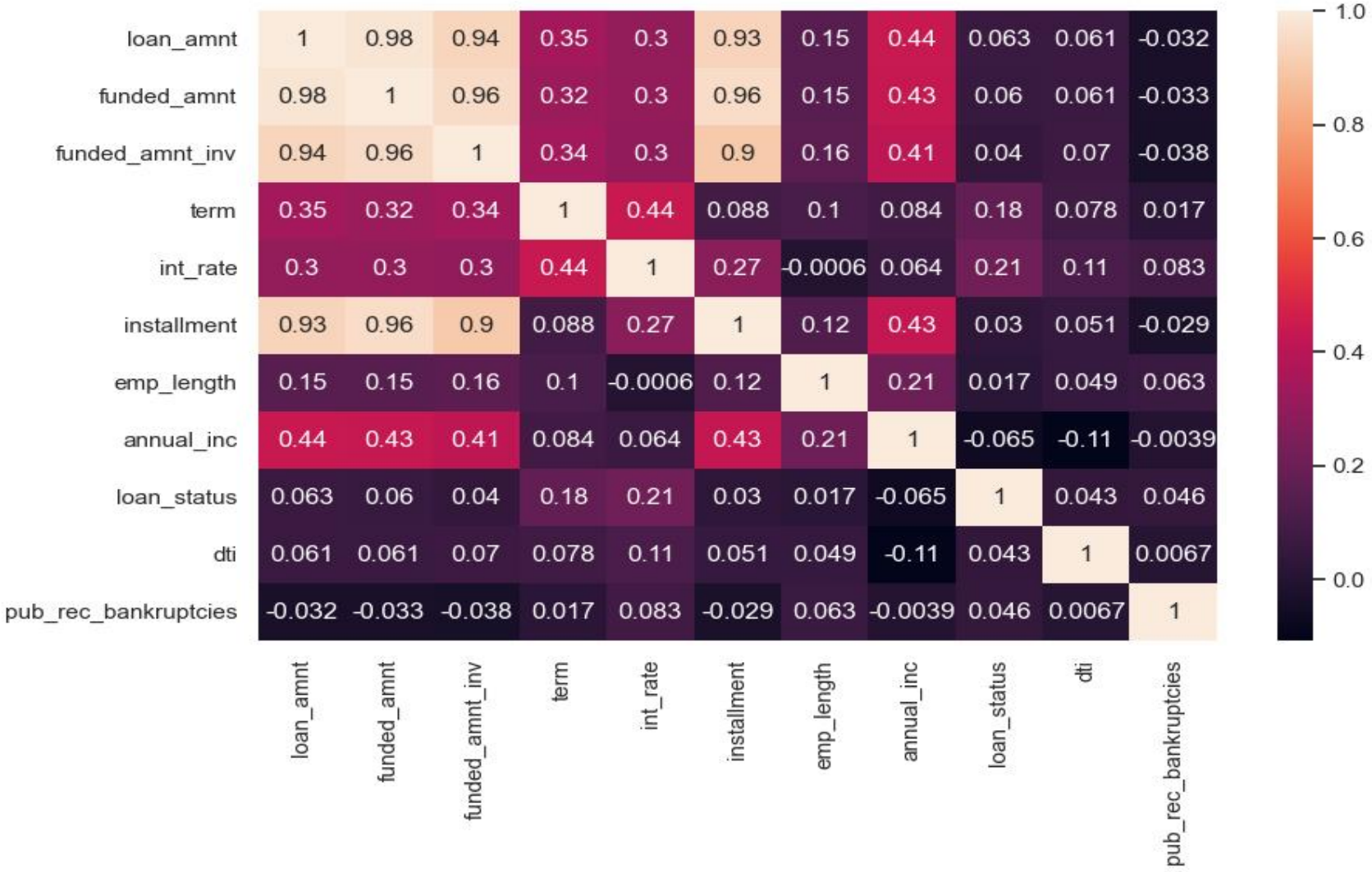
Most of the Defaulters fall in grade G.



More number of Charged Off loans falls under the verification status of not verified.



Correlation between variables

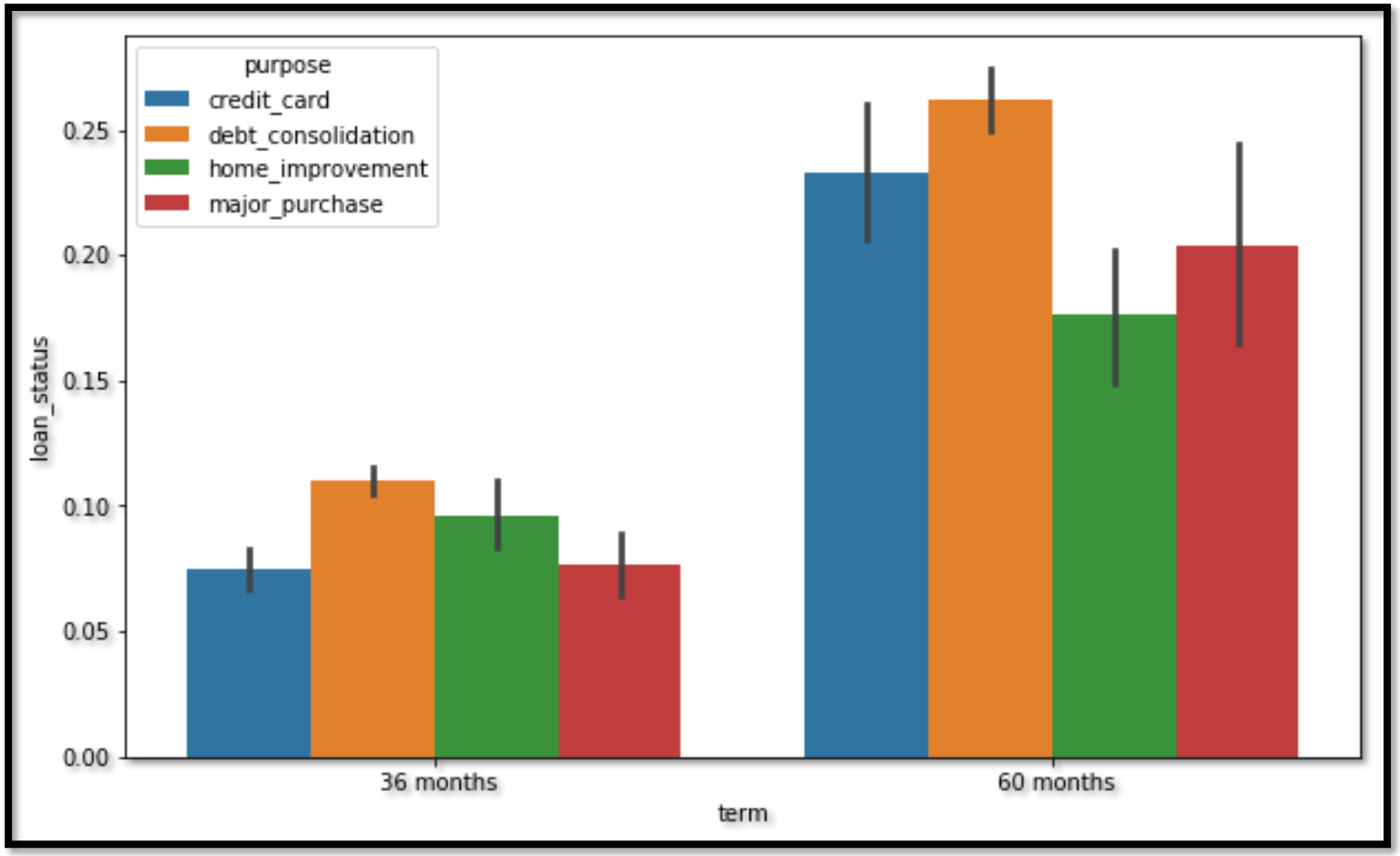


Highly Correlated

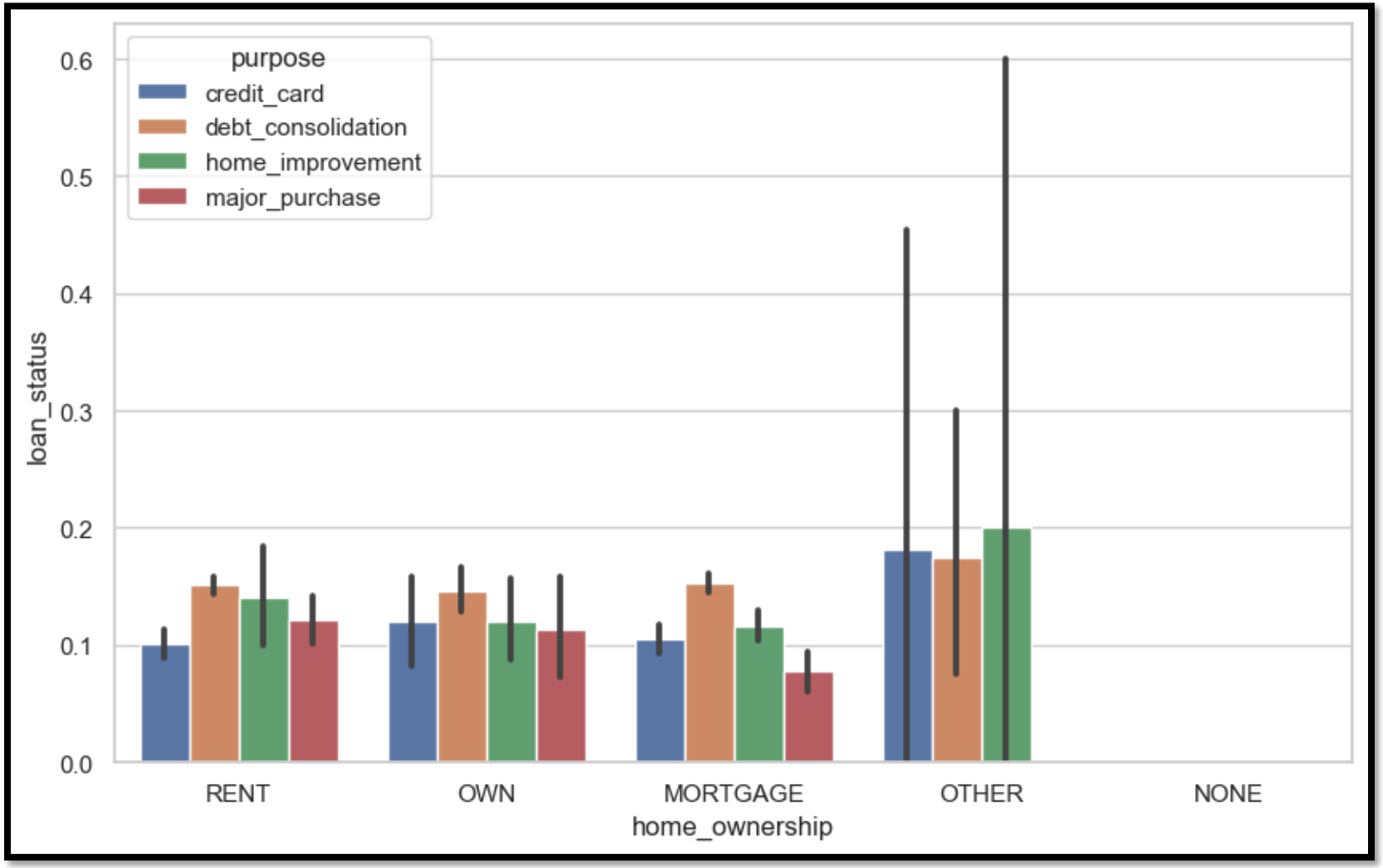
Loan_amnt	Funded_amnt
Loan_amnt	Funded_amnt_inv
Loan_amnt	Annual _inc
Loan_amnt	installment

Segmented Analysis

Debt consolidation loans have the highest default rates



Debt consolidation loans have the highest default rates



Variation of default rate across annual income

low	**0.16**
medium	**0.13**
very high	**0.10**
high	**0.10**

There is a 6% increase in default rate as we go from high to low annual income.